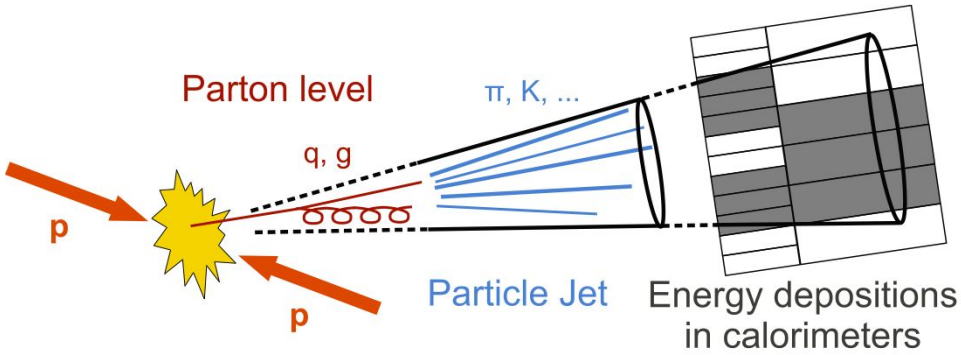
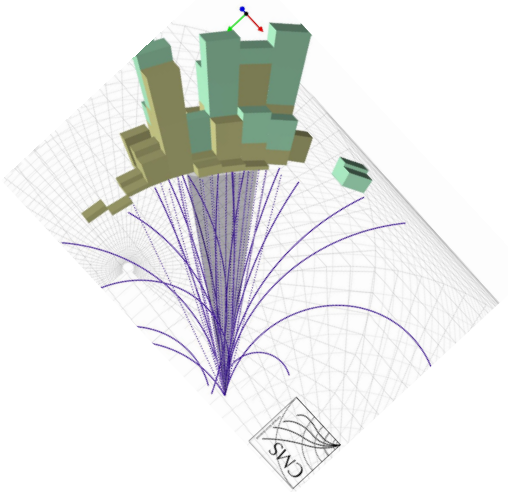
B-hadron identification in b-jets using novel deep learning technique in pp and PbPb collisions in CMS

Shamik Ghosh, Matthew Nguyen
(LLR Ecole Polytechnique - CNRS)

Jets in Colliders





- A lot of interesting physics at the LHC involves quarks
- quarks/gluons cannot exist on their own and hadronise
- Collimated sprays of particles in a cone are clustered and called jets
- Structure of these jets contain hints to their flavor type

Probing Jet Structure

The identification of jets originating from heavy-flavored particles is very important in CMS

- Study Higgs -> bb/cc (bb has > 50% BR)
- Interaction with medium following heavy-ion collision (dead-cone effect/ jet quenching)

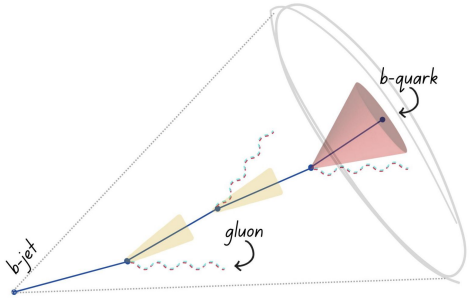
Important to identify b hadrons inside the b-jet

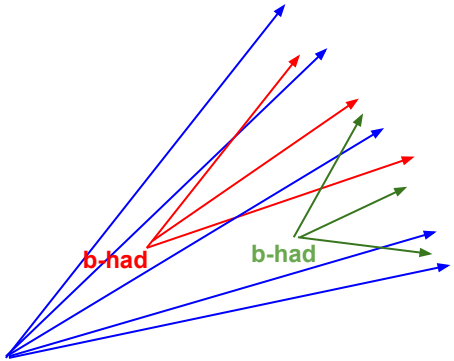
- General problem of estimating substructure of jets
  - Identification of boosted objects

Typical approach is to identify b-jet and then subsequently estimate properties

- CMS has excellent jet identification algorithms, ParticleNet and ParticleTransformer based

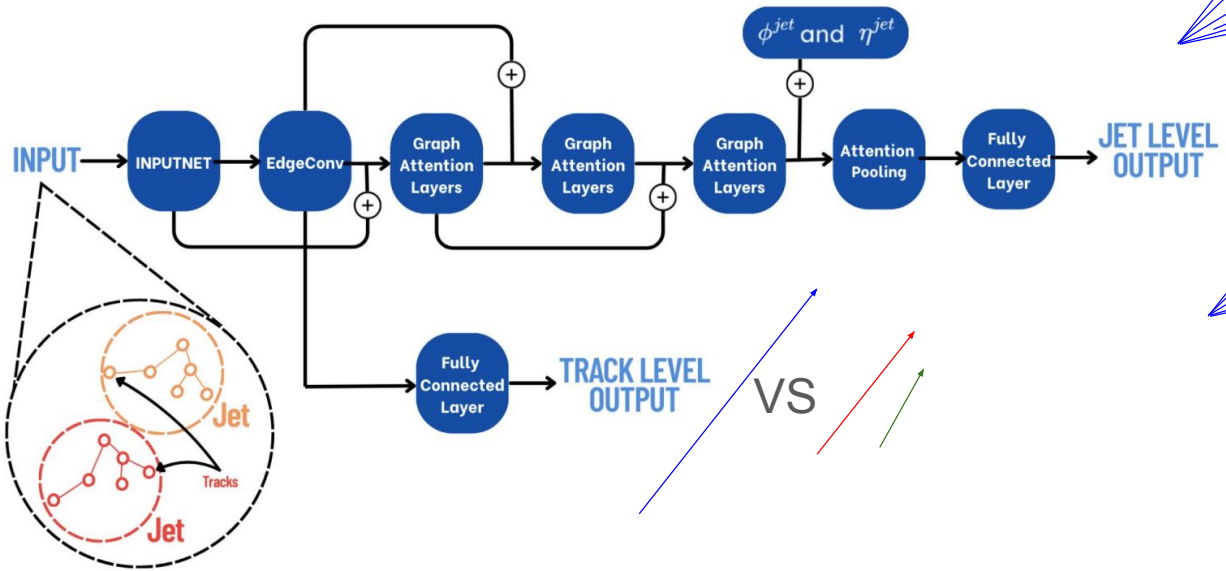
If done explicitly at the same time can lead to better performance for both





Identify b-hadron component in **red**/**green** from other jet components in **blue**

FusionNet: Design

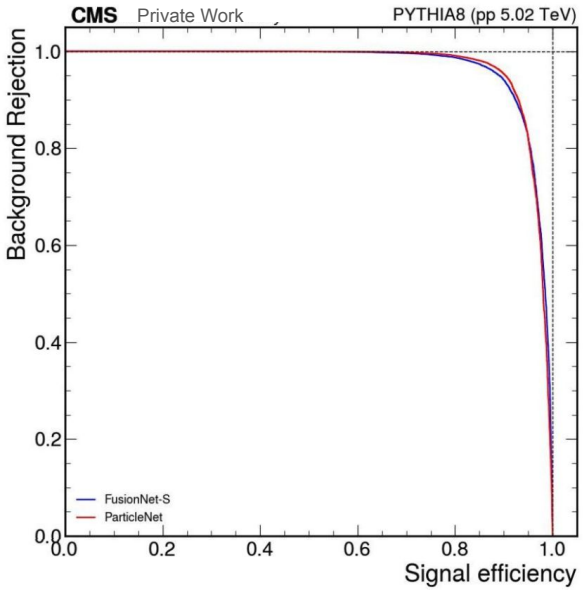


- Each charged particle is a node
- 16 node features: particle kinematic properties, track properties, secondary vertex properties
- Jet properties as global features
- Predict node-level and graph-level outputs

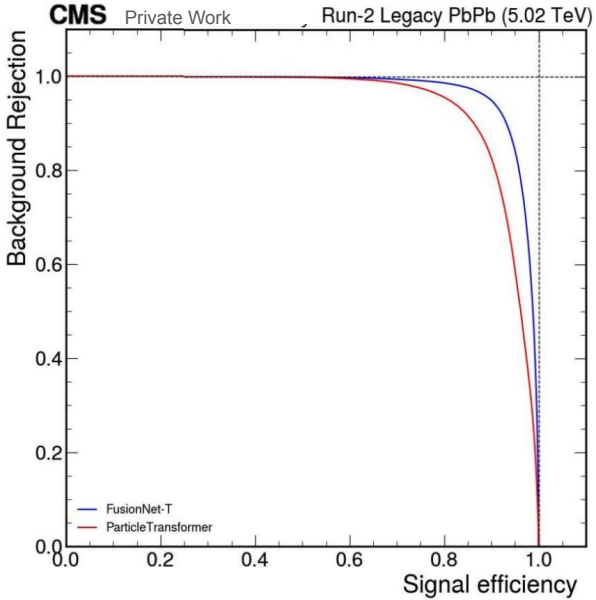
Training Details

- p-p 5.02 TeV : Number of jets: ~13M (Testing data: ~40K jets);
  Pb-Pb 5.02 TeV : Number of jets: ~5M (Testing data: ~140K jets)
- Truth labelling : Geometric ($\Delta R < 0.02$) and kinematic ($0.8 < p\_Track/p\_gen < 1.2$) matching
- Loss term : L = CrossEntropy(pred_jet, truth_jet) + CrossEntropy(pred_track, truth_trk)
- Learning rate annealing (Cosine)
- For Pb-Pb, model first pretrained on p-p

Results (Jet-Tagging)

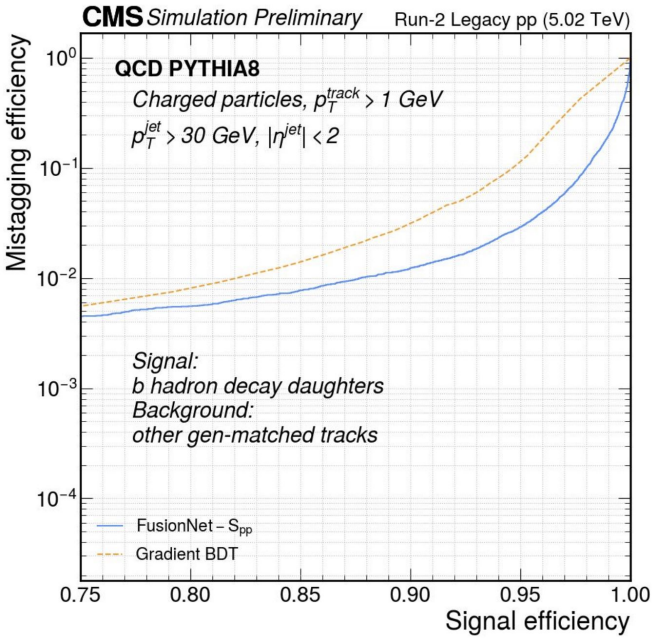


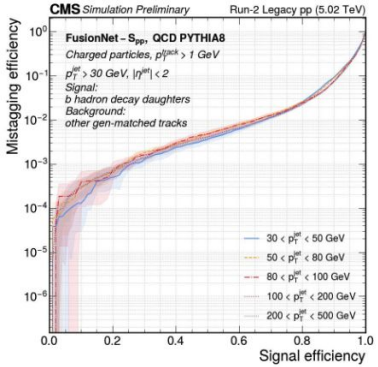
FusionNet performs as well as particlenet in p-p but uses only charged particle information

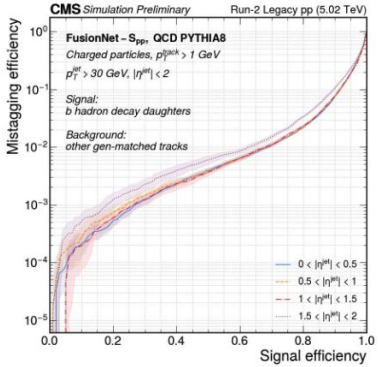


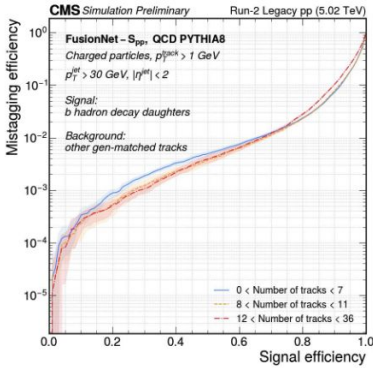
FusionNet shows performance improvement in Pb-Pb data

Results (Track-Tagging p-p)

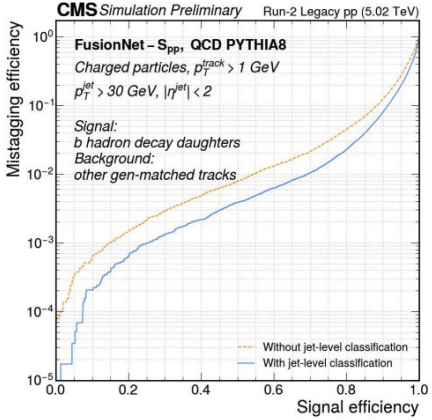








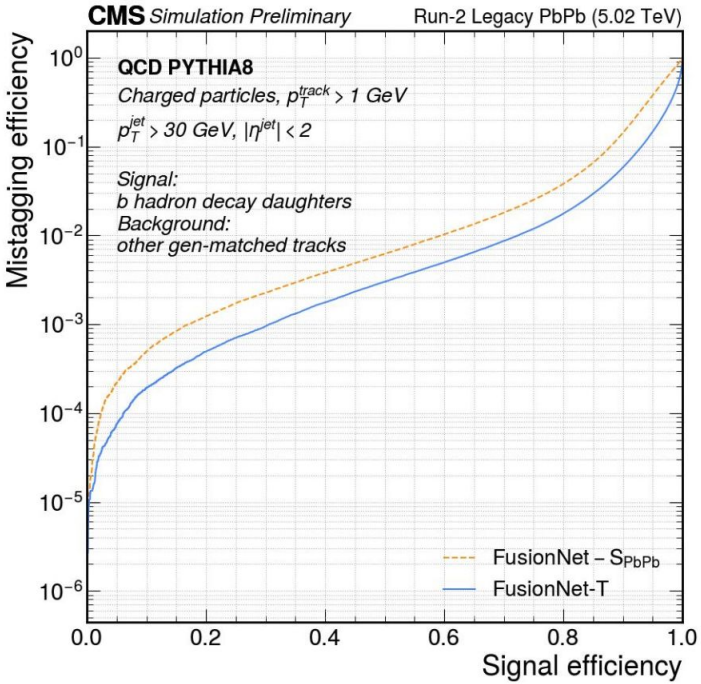
Performance of the novel model is stable



Joint training helps!

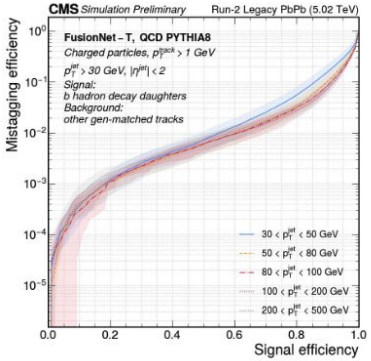
FusionNet significantly improves background suppression w.r.t existing BDT based tagger

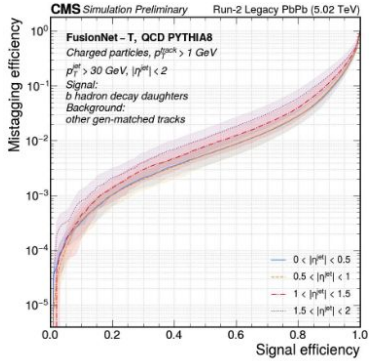
Results (Track-Tagging Pb-Pb)

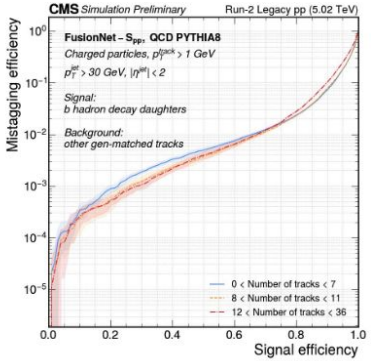


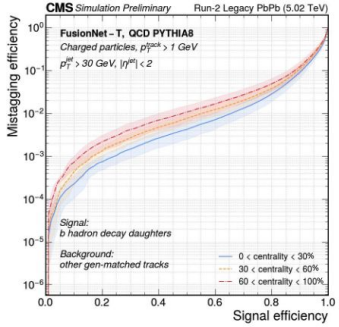
FusionNet provides SOAT performances

Performance improves upon pre-training on p-p









Performance of the novel model is stable

Summary

- ➔ FusionNet (CMS-DP-2025/035) jointly identifies b-jets and tracks from b-hadrons and achieves good performance in both
- ➔ Approach shows stable significant improvements over previous approaches
- ➔ Studies ongoing to extend this approach to other hadronic object tagging problems

Thanks for listening!

BACKUP